Privacy for the Average Citizen (April 2017)

Zach Morrissey

1. Abstract

As a general rule, private citizens in the US do not know how to properly safeguard their personal information online. Judging by the increase in volume and severity of data breaches over the past few years, it's becoming much harder for the average person to know how to conduct themself in a way that won't get their personal information in the hands of those who shouldn't have it. In addition to widespread data breaches, advertisers and other tracking software have become integrated parts of our online interactions, as a byproduct of the proliferation of social media. The aim of this vision document is to create a site for promoting privacy information and relevant services, so that the average layman can approach online security and know the best options available.

2. The Problem of Online Privacy

There are numerous recent examples of major private data breaches:

- Facebook allowing 3rd party vendors to take and resell individual users data.[1]
- Panera Bread leaked personal information and partial credit card numbers of a huge amount of its customers.[2]
- Exercise app MyFitnessPal forfeited the information to over 150 million accounts in a recent data breach.[3]

By some estimates, there were slightly less than 2 billion customer records leaked or stolen in the first half of 2017 alone.[4] Given that one's personal data online is always either being poorly guarded or monetized, how does one have an enjoyable experience online without descending into some sort of scam-ridden, advertisement-infested dystopia? There is a web out there that does not prey on your personal information for monetary gain, and the purpose of this project is to help users find their way to that web.

As a user, having data breaches and violations of public trust be a common thread in today's online world is a frustrating and powerless experience. It always follows the same cycle. First, you see news of a major privacy breach or data security issue. Next, there is inevitable fallout in the media and the company will release some sort of vapid statement of empathy for users and a less-than-stellar set of actions to try and cover for it. Then, as an end user, those who used their services are left out to dry and may have to fend with their personal information for purposes they may not want or agree with. In this part, there should be somewhere that people can easily take action in order to protect themselves.

3. Companies Do Not Keep User Data Safe By Default

It's not only end users that have difficulty with this; there are a number of reasons that companies do poorly at keeping users information safe. First, online security is a very difficult achievement. Successfully marketing any product today, regardless of whether it's a tech product or not, means that there will be a proliferation of data and services to accompany it. Companies don't have an infrastructure set up to properly address security concerns. For example, Troy Hunt of the website “Have I Been Pwned” contacted the toymaker CloudPets in December 2017 with information about how their site was easily compromised.[5] This meant that over 2.2 million recordings of extremely private information such as voice recordings of parents and their children playing with the CloudPets toys was available on a publicly web-accessible MongoDB that was not even secured with a password. This data was then deleted and held ransom by a malicious third party.[6] It was also revealed that others could send information to others’ toys, leading to scenarios where customers were unhappy and frightened to have this exposed.[7] In this scenario, we have a company that is likely not in any position to have major tech expertise available internally, and does not have any infrastructure for dealing with security threats.

Second, there is often very little penalty associated with having large amounts of user data collected, and a lack of consequences for keeping that information secure. Often, product manufacturers are in a race to the bottom concerning being the first and most affordable offering in the market, and consumers do not demand security and privacy as one of their primary requirements from products they buy.[8] These lead to scenarios like the current web-enabled security camera market, which has products so lacking in security that there are publicly available sites where you can watch others households being recorded live.[8] This lack of accountability is not solely for Internet of Things products either. Companies as large as Equifax, who leaked highly sensitive information for 143 million americans, have faced few consequences as a result of their actions. As of April 2018, over a year after the initial leaks occurred, no fines have been levied against the corporation nor has it lost its ability to stockpile user information without their consent.[9]

4. Online Security is a Growing Problem

The state of online privacy and data security is getting worse. There were 791 major data breaches cataloged in the first half of 2017, which is up 29% from the same period during the prior year.[10] In addition, the more recent breaches are higher in customer data record volume than prior breaches. The Equifax leak of mid-2017 released roughly 143 million records, and the 2013 breach of Yahoo released over 3 billion customer records.[10] Finally, as we learn about data breaches, it’s important to remember that often more information follows in the wake of the original revelations. This information can get lost in the noise surrounding the security disclosure. For example, in the fallout about Cambridge Analytica data leak, it became known that Facebook also scans the information that you send to others on Messenger.[11]

5. A Privacy Site to Help Users Stay Safe

The solution proposed here would be a website that catalogs and provides access to services that allow users to keep their information private. The service provides two purposes: writing guides that showcase technologies that can help users keep their data safe, and then have a place where users can rate/review individual services that provide those technologies. For example, there are multiple VPN providers out there, but not all of them are trustworthy or even help you keep your browsing private.

6. Service Features & Proposed Architecture

- The site would be a website written in HTML/CSS and Javascript for the interactive portions.
- The application server would be written in a web framework such as Flask or Express. The application itself would run on a Linux server.

- A Postgres/MySQL database that holds the data for the application server.
- For services that have some sort of stats that can be collected, a set of batch jobs that regularly track those and add that information to the database.
- Database queries that determine what are the top services as well as regular maintenance will be scheduled on the database server.
- If it turns out to be required, a caching layer in a key-value store such as Redis/Memcached could be used to reduced the amount of queries sent to the database and improve site performance.

For the database, schemas will be created to host the different types of privacy services available (VPN, browser extensions, adblockers, etc). These will be used to populate views and have users provide feedback on them. The webserver will capture request data, which will also be stored in the database, to determine user engagement metrics so that the most popular pages can be ranked and presented in an intuitive manner.

7. Key Requirements

The criteria for success in this project is that the users are able to use the page and gain valuable insight into how they can protect their information. For potential users of the site, this information should be made available in a way that there is no need to consult additional outside resources on the same subjects. Some more specific goals:

- Users should be able to authenticate to the server, and keep their information in an account on the page.
- Starting from the homepage, users can see and interact with content both from an educational standpoint (guides written by users) or a decision-making standpoint (reviews of services and information).
- For a specific guide, users can read through the information there and get enough context to be able to take action.
- For a specific service, users should be able to see the top-ranked services.
- For each of those services, users should be able to see feature lists, any relevant pricing information if they're not free, and statistics on how this service helps. Users should be able to read reviews written by other users who have familiarity with that offering.
- The website should either have a specific mobile-optimized version, or be responsive, so that reading/interacting on a phone is a seamless experience.

Keeping with our earlier example, a user may have written a guide to keeping your browsing private using a VPN, and then from there the users would browse reviews of VPN providers and select one for their own purposes.

8. Developing The Architecture

Fortunately, web development skills are very common, and the features required for a service like this are fairly straightforward to implement and have packages readily available. Some of the basic features, such as designing the data models for each type of service and implementing those in the database, are fairly straightforward. Others, such as authenticating users and keeping that information in the user session, are more difficult. All of the technologies used are free and openly available, so no licensing fees or product keys would be required to use them.

From the outset, designing the application server and writing the correct information to the database would take the most time. Between members of the team, this would likely take between 30 and 40 hours in total to implement a first proof-of-concept version, as well as another 10-20 hours to refine and implement feedback. For the final portion, I would expect that testing and user acceptance would take another 20 hours of work or so, to verify that the final product matches the users' vision.

9. Solving the Problem

First and foremost, the goal of the site is to be an accessible place where users can find information about how to protect themselves online. There is plenty of security information available, but much of it is geared towards users who are already technically savvy. The ideal audience for this site would be those who aren't tech savvy and could use the straightforward nature of a step-by-step guide on how to use them. By having a place where users can both find and provide feedback on solutions that are available, it will make it more natural to find ones that improve your life.

As the site becomes more popular, entrepreneurs will realize that there is a burgeoning market for data security for the individual. They will create services with the intent of sharing them on this platform, and this will begin a positive feedback loop of creating, sharing, and improving security for the end user.

10. Conclusion

This project is one that aims to make it simpler to get the right advice for keeping yourself secure on the internet. By providing a useful, interactive place where non-technical users can give and receive feedback on privacy and security services, there will ideally be a larger push for the layman to adopt privacy measures in their own life. The aim of the website is to specifically enable users who don't have much experience with online security, but buoyed by their frustration with widespread data security issues, are looking to protect themselves. This site will not fully solve the problem of online privacy and security, as even someone who has taken all of the protective measures can still be vulnerable, but the goal is to improve public perception and awareness of issues. In that, creating a useful tool with commonly known technologies like the web can go a long way in improving people's lives.

Citations:

1. K. Granville, "Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens," *The New York Times*, 19-Mar-2018. [Online]. Available: https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html. [Accessed: 05-Apr-2018].
2. "No, Panera Bread Doesnt Take Security Seriously – PB – Medium," *Medium*, 03-Apr-2018. [Online]. Available: https://medium.com/@djhoulihan/no-panera-bread-doesnt-take-security-seriously-bf078027f815. [Accessed: 05-Apr-2018].
3. N. Statt, "Under Armour says 150 million MyFitnessPal accounts compromised in data breach," *The Verge*, 29-Mar-2018. [Online]. Available: https://www.theverge.com/2018/3/29/17177848/under-armour-myfitnesspal-data-breach-150-million-accounts-security. [Accessed: 05-Apr-2018].

4. M. Leech, “Data breach statistics 2017: First half results are in,” *Gemalto blog*, 26-Sep-2017. [Online]. Available: https://blog.gemalto.com/security/2017/09/21/new-breach-level-index-findings-for-first-half-of-2017/. [Accessed: 05-Apr-2018].
5. T. Hunt, “Data breach disclosure 101: How to succeed after you've failed,” *Troy Hunt*, 25-Mar-2017. [Online]. Available: https://www.troyhunt.com/data-breach-disclosure-101-how-to-succeed-after-youve-failed/. [Accessed: 01-Apr-2018].
6. T. Hunt, “Data from connected CloudPets teddy bears leaked and ransomed, exposing kids' voice messages,” *Troy Hunt*, 20-Dec-2017. [Online]. Available: https://www.troyhunt.com/data-from-connected-cloudpets-teddy-bears-leaked-and-ransomed-exposing-kids-voice-messages/. [Accessed: 05-Apr-2018].
7. R. Chirgwin, “CloudPets woes worsen: Webpages can turn kids stuffed toys into creepy audio bugs,” *The Register® - Biting the hand that feeds IT*. [Online]. Available: https://www.theregister.co.uk/2017/03/01/cloudpets_woes_worsen_mics_can_be_pwned/. [Accessed: 05-Apr-2018].
8. J. M. Porup, “‘Internet of Things’ security is hilariously broken and getting worse,” *Ars Technica*, 23-Jan-2016. [Online]. Available: https://arstechnica.com/information-technology/2016/01/how-to-search-the-internet-of-things-for-photos-of-sleeping-babies/. [Accessed: 04-Apr-2018].
9. J. Puzzanghera, “Senators want 'massive' fines for data breaches at Equifax and other credit reporting firms,” *Los Angeles Times*, 10-Jan-2018. [Online]. Available: http://www.latimes.com/business/la-fi-equifax-data-breach-fines-20180110-story.html. [Accessed: 03-Apr-2018].
10. G. Bloom, “Why Data Breaches Will Get Worse Before Things Get Better,” *Forbes*, 29-Nov-2017. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2017/11/29/why-data-breaches-will-get-worse-before-things-get-better/#74ffb101339f. [Accessed: 02-Apr-2018].
11. S. Frier, “Facebook Scans the Photos and Links You Send on Messenger,” *Bloomberg.com*, 04-Apr-2018. [Online]. Available: https://www.bloomberg.com/news/articles/2018-04-04/facebook-scans-what-you-send-to-other-people-on-messenger-app. [Accessed: 02-Apr-2018].